我国寿险业粗死亡率的实证研究*

王艺明 陈浪南

ABSTRACT

The paper conducts experimental study of crude death rate of insurance industry by using turning-point-having Logistic Model.

关键词: 粗死亡率; Logistic 模型; 转折点

一、绪论

死亡率是人类面临死亡风险的概率,而生命表则是死亡率按年龄归纳成的统计表。目前生命表被广泛应用于保险业,成为各保险公司计算保险费和责任保证金的基础。生命表中各年龄级别的死亡率是否能真正反映保单死亡率,直接关系到寿险公司是否能计算出合理的保险费率并正常地运营,以及保户缴交保险费的高低。目前我国寿险公司计算死亡率的基础为 1990～1993 年中国人寿保险业经验生命表(China Life Insurance Mortality Table、1990～1993),至今已有近 10 年。由张揖平等(1998)的分析可知:寿险业的粗死亡率有逐年下降的趋势,这可能是因为卫生条件改善、医疗技术进步、个人对健康的重视及核保技术提高等因素所造成。然而他们使用 Logistic 模型所预估的粗死亡率在高年龄层有低估的现象。

本文的主要目的是应用具有转折点的 Logistic 模型,即在转折点前后参数不同的 Logistic 模型,对我国经验生命表所反映的保单粗死亡率作实证研究。目前,国内该领域的研究尚属空白。虽然,转折点的模型在文献上已有很多探讨,但是多数局限于以下两种情形:一是假设独立同分布情况下,如 Hinkley & Hinkley(1970) 假设二项式分布,Hinkley(1970) 假设正态分布,另一种是假设传统的回归模型情况下。至于具有转折点的 Logistic 模型的讨论,目前仅见于张揖平(1999)。

本文将以中国人寿保险业经验生命表(1990～1993)中,保险年龄为 15 岁至 74 岁男性和女性的保险死亡人数(actual deaths)、暴露数(exposed to risk)和粗死亡率(crude mortality rate)等数据为研究对象,应用具有转折点的 Logistic 模型作实证分析。

二、研究模型与方法

张揖平(1998)发现无论男性或女性的死亡率均大约在 15 岁以后和年龄成一非线性递增关系,虽然许多分析死亡率的研究均采用修匀(gradtlation)技巧,可参考 London(1985)的讨论,但大部分的修匀技巧均强调数值计算和平滑(smooth)的假设,有关参数模型优于一般着重平均修匀

*本文为国家社科基金课题(03AJY008)和厦门大学校级课题的成果之一。

5. 管理风险

从事期货避险交易时可能牵涉新的资源与人力配置以及相关的报表流程、会计、税务与法规等,加上国内投资者在刚刚推出指数期货交易时不太熟悉期货交易,管理者若无法有效规范与监控整个流程可能会产生管理风险。

参考文献

[1] Charles M. S, Stock Index Futures: Theories and International Evidence [M]. Second Edition, International Thomson Business Press, 1997.

[2] John C. Hill, Options, Futures and other Derivative Securities[M]. Prentice-Hall, Inc, 1993.

作者簡介

徐国祥,经济学博士,上海财经大学应用统计研究中心主任、统计学系教授、博士生导师,上海市统计学会副会长,上海财经大学会计与财务研究院研究员。

檀向球,经济学博士,申银万国证券研究所研究员。

(责任编辑:何平)

方法的原因可参考 Shaban(1980)。为此,本文在比较分析的过程中采用的参数模型为有转折点的 Logistic 模型。

(一)研究模型

令反应变量(response variable) Y_i 服从二项式分布,参数为 n_i 和 p_i,$i=1,2,\cdots,m$,其中 Y_i 表示死亡人数,n_i 表示暴露数,p_i 表示平均粗死亡率,Y_i/n_n 即表示粗死亡率。本文主要探讨 1990～1993 年中国寿险业男性与女性粗死亡率的预估模型,由张揖平(1998)可知:若使用 Logistic 模型预估粗死亡率,虽然大部分年龄层均预估不错,但在高年龄层均有低估的现象,因此本文采用具有转折点的 Logistic 模型解决该问题,也即转折点前后参数不同的 Logistic 模型。假设影响粗死亡率的解释变量(explanation variable) x_i 表示年龄,τ 为转折点参数,d_i 表示虚拟变量(dummy variable),令

$$d_i=\begin{cases}0 & \text{如果 } x_i\le\tau\\ 1 & \text{如果 } x_i>\tau\end{cases}\quad(1)$$

其 Logistic 模型为:

$$\begin{aligned}\log it(p_1)&=\log(p_i/(1-p_i))\\&=\beta_0+\beta_1(x_i-\tau)+\beta_2(x_i-\tau)d_i,\end{aligned}\quad(2)$$

也即当年龄小于或等于 τ 时,斜率为 β_1,当年龄大于 τ 时,斜率为 β_2,且当年龄为 τ 时,$\log it(p_i)=\beta_0$。因此式(2)可视为连续转折情况。式(2)经过计算,p_i 可表示为:

$$p_i=\frac{\exp(\beta_0+\beta_1(x_i-\tau)+\beta_2(x_i-\tau)d_i)}{1+\exp(\beta_0+\beta_1(x_i-\tau)+\beta_2(x_i-\tau)d_i)}\quad(3)$$

其中 τ、β_0、β_1 和 β_2 均为未知参数。

(二)检验假设与方法

在具有转折点的 Logistic 模型下,Y_1 服从二项式分布,参数为 n_i 和 p_i,$i=1,2,\cdots,m$,其中 p_i 满足式(2),当 $\beta_2=0$ 表示转折点不存在,当 $\beta_2\neq0$ 则表示转折点确实存在。如果要检验是否有转折点存在,考虑

$$H_0:\beta_2=0\qquad v.s.\qquad H_1:\beta_2\neq0$$

本文采用两种检验方法:改进的似然比检验(modification of the likelihood ratio test)和改进的沃尔德检验(modified of the wald's test)。

1. 改进的似然比检验

若 H_1 成立,则似然函数(likelihood function)为

$$L_{H1}(\tau,\beta_0,\beta_1,\beta_2)=\prod_{i=1}^{m}(n_i\quad y_i)^T p_i^{y_i}(1-p_i)^{n_i-y_i}\quad(4)$$

因此

$$\begin{aligned}&\log H_{H1}(\tau,\beta_0,\beta_1,\beta_2)\\&=\sum_{i=1}^{m}\left\{\log(n_i\quad y_i)^T+y_i\log(p_i)+(n_i-y_i)\log(1-p_i)\right\}\\&=\sum_{i=1}^{m}\left\{\log(n_i\quad y_i)^T+y_i\log(p_i/(1-p_i))+n_i\log(1-p_i)\right\}\\&=\sum_{i=1}^{m}\left\{\log(n_i\quad y_i)^T+y_i\eta_i+n_i\log(1+\exp(\eta_i))\right\}\end{aligned}\quad(5)$$

其中 $\eta_i=\beta_0+\beta_1(x_i-\tau)+\beta_2(x_i-\tau)d_i\quad(6)$

若 H_0 成立,则

$$\begin{aligned}\log L_{H0}(\beta_0,\beta_1)=\sum_{i=1}^{m}\Big\{&\log\begin{pmatrix}n_i\\y_i\end{pmatrix}+y_i(\beta_0+\beta_1(x_i-\tau))\\&+n_i\log(1+\exp(\beta_0+\beta_1(x_i-\tau)))\Big\}\end{aligned}\quad(7)$$

因此,如果假设转折点 τ 已知,则似然比检验(likelihood ratio test)为

$$\begin{aligned}\lambda(\tau)&=\max_{\beta_0,\beta_1,\beta_2}\log L_{H1}(\tau,\beta_0,\beta_1,\beta_2)\\&\quad-\max_{\beta_0,\beta_1}\log L_{H0}(\beta_0,\beta_0)\end{aligned}\quad(8)$$

其中 $\max_{\beta_0,\beta_1,\beta_2}\log L_{H1}(\tau,\beta_0,\beta_1,\beta_2)$ 和 $\max_{\beta_0,\beta_1}\log L_{H0}(\beta_0,\beta_1)$ 可经 Fisher-scoring 的迭代(iterative)方法求得。当 H_0 成立时,$2\lambda(\tau)$ 其分布近似 $\chi^2(1)$,因此,如果 α 为显著性水平,当

$$2\lambda(\tau)\ge\chi_\alpha^1(1)\quad(9)$$

成立时,表示有足够证据拒绝 H_0,其中 $\chi_\alpha^1(1)$ 为 $\chi^2(1)$ 右尾的第 100α 百分位。如果转折点 τ 为未知,则改进的似然比检验统计量为

$$\begin{aligned}&\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda(\tau)\\&=\max_{\tau_0+1\le\tau\le m-\tau_0}\max_{\beta_0,\beta_1,\beta_2}\log L_{H1}(\tau,\beta_0,\beta_1,\beta_2)\\&\quad-\max_{\beta_0,\beta_1}\log L_{H0}(\beta_0,\beta_1)\end{aligned}\quad(10)$$

其中 τ_0 一般取为 0.05m,而只在 τ_0+1 和 $m-\tau_0$ 中寻找 $\lambda(\tau)$ 的极大值,其主要目的是希望转折点不会发生于边界附近。当 $2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda(\tau)$ 足够大时,表示有足够证据拒绝 H_0。可应用 Boole 不等式求 $2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda(\tau)$ 的 p-value 的上界,p-value 定义为

$$P_{H0}(2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda(\tau)\ge2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda_{obs}(\tau))\quad(11)$$

其中 $\lambda_{obs}(\tau)$ 表示将观察值代入 $\lambda(\tau)$ 中,根据 Boole 不等式,有以下结果

$$\begin{aligned}&P_{H0}(2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda(\tau)\ge2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda_{obs}(\tau))\\&=P_{H0}\left\{\bigcup_{i=\tau_0+1}^{m-\tau_0}(2\lambda(i)>2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda_{obs}(\tau))\right\}\\&\le\sum_{i=\tau_0+1}^{m-\tau_0}P_{H0}(2\lambda(i)\ge2\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda_{obs}(\tau))\\&\le\sum_{i=\tau_0+1}^{m-\tau_0}P_{H0}(2\lambda(i)\ge2\lambda_{obs}(i))\equiv P_U\end{aligned}\quad(12)$$

如果 P_U 小于给定的显著水平 α 则拒绝 H_0。

2. 改进的沃尔德检验

如果转折点 τ 假设为已知,其沃尔德检验统计量为

$$W(\tau)=\left\{\beta_2(\tau)^2\right\}/\hat{V}ar(\beta_2(\tau))\quad(13)$$

其中 $\beta_2(\tau)$ 为 β_2 的最大似然估计(maximum likelihood estimator),$\hat{V}ar(\beta_2(\tau))$ 为 $\beta_2(\tau)$ 的方差的估计量。当 H_0 成立时,$W(\tau)$ 的分布近似 $\chi^2(1)$,因此,若 α 为显著水平,当

$$W(\tau)\ge\chi_\alpha^1(1)\quad(14)$$

成立时,表示有足够证据拒绝 H_0,其详细结果可参考 SAS(1996)。若转折点 τ 为未知,其改进的沃尔德检验统

计量为

$$\max_{\tau_0+1\le\tau\le m-\tau_0} W(\tau) \tag{15}$$

τ_0 也取为 $0.05m$。而其 p-value 也可用类似改进的似然比检验法做出,由 Boole 不等式求出其 p-value 的上界作为判断标准。

(三)参数估计

当 β_2 显著不为 0 时,也即转折点确实存在,因此需要估计 τ 和其余参数 β_0、β_1 和 β_2。如果转折点 τ 假设为已知,β_0、β_1 和 β_2 的估计量 $\beta_0(\tau)$、$\beta_1(\tau)$ 和 $\beta_2(\tau)$ 可以利用最大似然法求得,也即

$$\log L_{H1}(\tau,\beta_0(\tau),\beta_1(\tau),\beta_2(\tau))$$

$$=\max_{\beta_0,\beta_1,\beta_2}\log L_{H1}(\tau,\beta_0,\beta_1,\beta_2) \tag{16}$$

如果转折点 τ 为未知,可先利用检验的结果求得 τ 的估计量,例如,改进的最大似然比检验法为求 τ_L 使得

$$\lambda(\tau_L)=\max_{\tau_0+1\le\tau\le m-\tau_0}\lambda(\tau) \tag{17}$$

其中 $\lambda(\tau)$ 定义于式(8)。因此 β_0,β_1 和 β_2 的估计量可定义为 $\beta_0(\tau_L)$、$\beta_1(\tau_L)$ 和 $\beta_2(\tau_L)$。而改进的沃尔德检验法为求 τ_W 使得

$$W(\tau_W)=\max_{\tau_0+1\le\tau\le m-\tau_0} W(r) \tag{18}$$

其中 $W(r)$ 定义于式(13),因此 β_0,β_1 和 β_2 的估计量可定义为 $\beta_0(\tau_W)$、$\beta_1(\tau_W)$ 和 $\beta_2(\tau_W)$。

三、实证结果与分析

我们主要应用具有转折点的 Logistic 模型对我国 1990～1993 观察年度 15 至 74 岁男性和女性的粗死亡率数据作实证分析,应用上文介绍的方法,根据下列步骤:①取 $m=60$,$\tau_0=0.05m=3$;②对 $\tau=4,\cdots,57$,分别计算 $\lambda(\tau)$ 和 $W(\tau)$,其中 $\lambda(\tau)$ 和 $W(\tau)$ 分别定义于式(8)和(13),并且求得对应的 p-value,该部分可以通过 SAS 软件计算得到;③通过比较得到的 $\max_{4\le\tau\le57}\lambda(\tau)$ 和 $\max_{4\le\tau\le57}W(\tau)$,再由式(12),可求得检验统计量 $\max_{4\le\tau\le57}\lambda(\tau)$ 的 p-value 的上界;以同样方法可求得检验统计量 $\max_{4\le\tau\le57}W(\tau)$ 的 p-value 的上界;④显著水平 α 取为 0.01,若 p-value 的上界小于 0.01,则有足够证据显示 β_2 不为 0,也即转折点显著存在,并且由式(17)和(18)可求得转折点发生在何处。

经过计算可以得到:就男性而言,使用改进的似然比检验,检验统计量 $\max_{4\le\tau\le57}\lambda(\tau)=563.54$,其 p-value 的上界小于 0.01,使用改进的沃尔德检验,检验统计量 $\max_{4\le\tau\le57}W(\tau)=427.05$,其 p-value 的上界也小于 0.01;就女性而言,使用改进的似然比检验,检验统计量 $\max_{4\le\tau\le57}\lambda(\tau)=173.22$,其 p-value 的上界小于 0.01,使用改进的沃尔德检验,其检验统计量 $\max_{4\le\tau\le57}W(\tau)=183.60$,其 p-value 的上界也小于 0.01。因此无论男性或女性的粗死亡率均有显著转折情况发生。而利用上文的参数估计方法估计未知参数其结果如表 1,其中 MLRT 表示由改进的似然比检验法求出的估计值,MWT 表示由改进的沃尔德检验法求出的估计值,由此可知男性的粗死亡率大约在 31～32 岁左右有一个明显的转折点发生,而女性的粗死亡率大约在 28～29 岁左右有一个明显的转折点发生。

表 1 不同方法的参数估计

	男性		女性	
	MLRT	MWT	MLRT	MWT
τ	32	31	29	28
β_0	-7.7898	-7.7898	-6.6231	-7.8715
β_0	-0.0029	-0.0029	0.0058	-0.0375
β_0	0.0930	0.0930	0.0930	0.1061

四、结论与今后研究方向

本文主要应用具有转折点的 Logistic 模型建立 1990～1993 观察年度中国 15 至 74 岁男性和女性的粗死亡率模型,并应用改进的似然比检验法和改进的沃尔德检验法检验是否有转折点,如果有转折点发生则估计转折点发生在何处,并求出其余未知参数,用以建立粗死亡率模型。结论为男性的粗死亡率大约在 31～32 岁左右有一个明显的转折点发生,而女性的粗死亡率大约在 28～29 岁左右有一个明显的转折点发生。

就粗死亡率而言,本文假设至多只有一个转折点,文献上虽有一些研究假设不止一个转折点,甚至转折点个数未知,其作法大致为先假设至多一个转折点,如果检验出有显著转折点时,再由转折点前后分为两段数据,分别讨论是否再有转折情况发生,重复该步骤,直到每段数据都显示无转折情况(Srivastava & Worsley,1986),但文献上均是在一些较简单的模型下讨论,如果要推广到具有转折点的 Logistic 模型将是重要而且复杂的问题,可作为进一步研究的方向。

参考文献

[1] Hinkley, D. V. (1970), "Inference about the Change-Point in a Sequence of Random Variables," Biometrika, 57, pp. 1～17.

[2] Hinkley, D. V. and Hinkley, E. V. (1970). "Inference about the Change-Point in a Sequence of Binomial Variables," Biometrika, 57, pp. 477～488.

[3] London, D. (1985), Graduation: The Revision of Estimates, Actex, Connecticut.

作者简介

王艺明,陈浪南,厦门大学、中山大学岭南学院经济研究所。

(责任编辑:石庆焱)